Artificial Intelligence for Speech Recognition Based on Neural Networks

Takialddin Al Smadi[1], Huthaifa A. Al Issa[2], Esam Trad[3], Khalid A. Al Smadi[4]

[1]Department of Communications and Electronics Engineering, College of Engineering, Jerash University, Jerash, Jordan
[2]Department of Electrical and Electronics Engineering, Faculty of Engineering, Al-Balqa Applied University, Al-Huson College University, Al-Huson, Jordan
[3]Departments of Communications and Computer Engineering, Jadara University, Irbid, Jordan
[4]Jordanian Sudanese Colleges for Science & Technology, Khartoum, Sudan
Email: dsmadi@rambler.ru

Abstract

Speech recognition or speech to text includes capturing and digitizing the sound waves, transformation of basic linguistic units or phonemes, constructing words from phonemes and contextually analyzing the words to ensure the correct spelling of words that sounds the same. Approach: Studying the possibility of designing a software system using one of the techniques of artificial intelligence applications neuron networks where this system is able to distinguish the sound signals and neural networks of irregular users. Fixed weights are trained on those forms first and then the system gives the output match for each of these formats and high speed. The proposed neural network study is based on solutions of speech recognition tasks, detecting signals using angular modulation and detection of modulated techniques.

1. Introduction

Artificial intelligence applications have proliferated in recent years, especially in the applications of neural networks where they represent an appropriate tool to solve many problems highlighted by distinguished styles and classification.

The year of 1943 is known as the beginning of the evolution of artificial neural systems.

The first formal model of neurons through a computer model that includes all the necessary elements and the completion and implementation of the electronic form of this model is not practical or reasonable in terms of tech during the vacuum tube. It should be noted that this model has been applied extensively to describe computer hardware for the vacuum tube [1]. Initially, planned tutorial to update connections of nerve cells that are referred to the law educational learning rule HYIP has stated that the information can be stored in the links and connections. It is recognized that learning technology has proved its benefits in the future development of this field. Hip education Act initial contribution in neural network theory had been built and tested in the first study of the neurological computer in the 1950s, where the application contacts automatically and during this stage the term preceptor called the unit represented for neural cell to invent the term world and divorced on the neuron, he pioneered the term frank Rosenblatt in 1958. This invention was a viable training machine learning and classification of certain models by modulating communication components first. In this way it has become along with the imagination of engineers and scientists and a background to the calculations of this type of machinery which is still used today.

In the early 1960s, a new created method called Adaptive Linear Combiner developed a very useful law [2].

2. Pattern Recognition

Automatic recognition, description, classification and grouping patterns are important parameters in various engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence and remote sensing. The template can be fingerprint images, handwritten words cursive, a human face or the voice signal. Given the pattern, its recognition/classification may be one of the following two tasks [3].
a) Under the supervision of a classification, discriminated analysis, in which the input pattern is defined as a member of a predefined class;
b) Unsupervised classification, clustering in which is the class template is unknown.

Recognition of the problem here is as a classification or classification problems, where the classes are defined by either the system designer in a controlled classification or learned based on similar models in unsupervised classification.

These applications include data mining the definition of "plan". For example, he correlations or independently in millions of multidimensional models, document classification effectively search text documents, financial, forecasting, organization and retrieval of multimedia databases and biometrics. The rapidly growing and available computing power, enabling faster processing of huge amounts of data, also promoted the use of complex and diverse methods for classification and analysis of data. At the same time, the demand for automatic pattern recognition is growing due to the presence of large databases and strict requirements speed, accuracy and cost. Design of recognition system template essentially consists of the following three aspects:
a) Collection and preprocessing, data reporting;
b) Decision-making process;
c) Scope dictates the choice of pretreatment technique.

Schema view and decision making models It is recognized that the problem of clearly defined and sufficiently limited recognition will lead to the introduction of the compact model and simple decision-making strategy. Learning from a set of examples is an important and necessary attribute of most systems of recognition template.

The most prominent approaches for pattern recognition are:
a) Matching pattern;
b) Statistical classification;
c) Syntactic or structural conformity and neural networks.

3. Neural Networks

Neural networks consist of a set of nodes that a special type of account collectively and that each node is the standard unit of account and the contract could work in parallel depends on the interactions among themselves and how they relate to some of the scholars are defined as:
• Mathematical models simulating characteristics of biological systems that deal with information in parallel composed of relatively simple elements called.
• Is a simple entity class of algorithms that are formulated in charts (graphs grouped these schemes a large num-

ber of algorithms and these algorithms provide solutions to a number of complex problems [4].

To highlight the activity of neural networks is the process of classification and coding and to highlight the properties of neural networks are:

a) Resistance to noise;
b) Flexibility in dealing with the distorted images;
c) Maximum resistance to tag images of dismembered or partially decomposed;
d) Combinations of parallel processes with a large number of operating units that stimulate by interdependence of processes in addition to the stock of information distributed in parallel.

With non-linear operations, *i.e.* their ability to make non-linear relationships include maps of noise that makes them a good source of ratings and attribution (classification predication);

e) High capacity to adapt the system of logarithms and powers of education internal allows the use of internal adjustment that lives in the vicinity of lasting change.

Types of Neural Networks

Possible to identify the most common types of neural networks with input types and learn some common uses as in **Table 1** shown [5] [6].

4. Procedure Works

The method consists of iteratively selecting the most distant score with respect to mean. If this score goes beyond a certain threshold, the score is removed and mean and standard deviation estimations are recalculated. When there are only a few utterances to estimate mean and variance, this method leads to a great improvement. Text dependent and text independent experiments have been carried out by using a telephonic multisession database. The paper presents the inter-relationship between algorithmic research system developments based on the experience from the speaker using mini-problems during the system design process, and presents a model of speech recognition based on artificial neural networks [7]. **Figure 1** shows the diagram of the processing of speech signals.
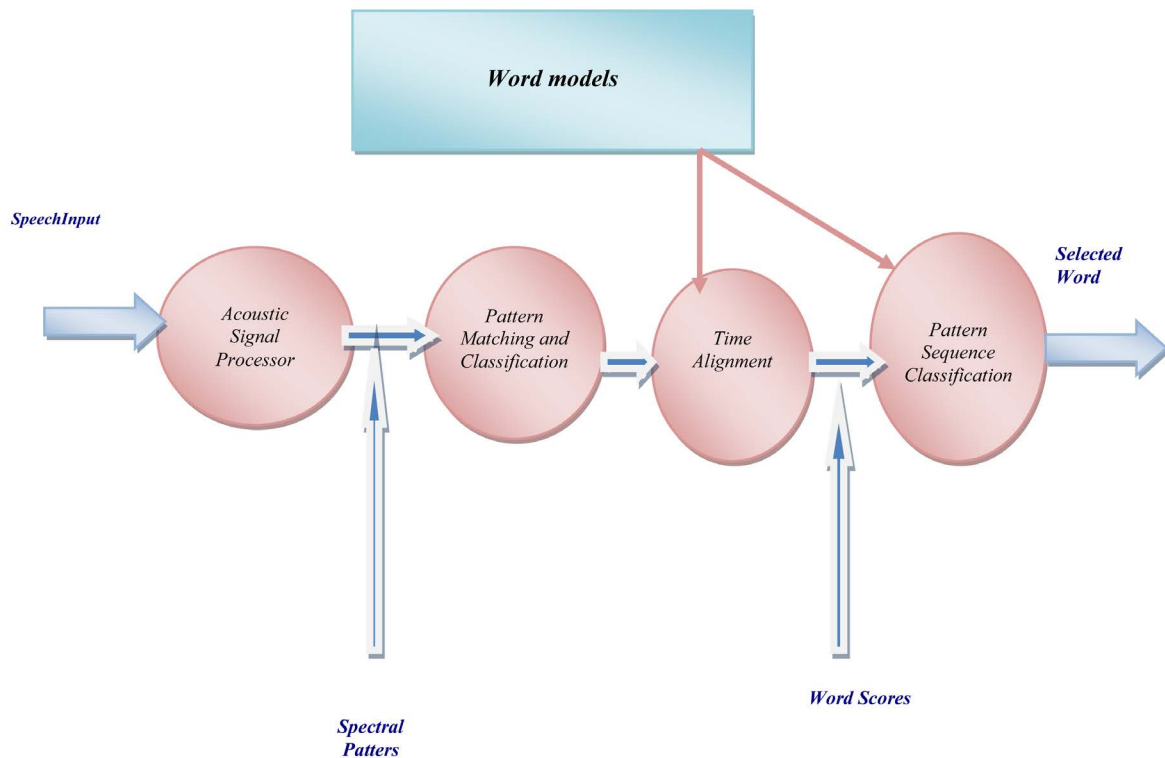

Figure 1. Diagram of the processing of speech signals planning.

Table 1. Types of neural networks and application.

Common uses	Input method	Input type	Types of neural networks
Associated memory to distinguish ASCII characters	Supervised	Binary	**Hopfield-Net**
Connect with similar dual channel	Supervised	Binary	**Hammin_Net**
Assembly (adaptive resonance theory)	Supervised	Binary	**Carpenter/grassbery classifier**
Discrimination and classification of simple shapes	Supervised	Continuous	**Perceptron**
Featuring complex shapes and classification	Supervised	Continuous	**Multi-layer perceptron**
Evaluation of vector and speech, and analogy to biological neural networks	**Supervised**	**Continuous**	**Kohonenself organizing feature map**

a) Present study of artificial neural networks for speech recognition task. Neural network size influence on the effectiveness of detection of phonemes in words. The research methods of speech signal parameterization. Learn about how to use linear prediction analysis, a temporary way of learning of the neural network for recognition of phonemes. The proposed way of teaching as input requires only the transcription of words from the training set and do not require any manual segmentation of words;

b) Development and research of the methods for diagnosing and detecting modulated signals;

c) Software implementation and pilot testing on real signals of neural network methods for processing.

4.1. Recognition Process Recognition Algorithm

- Input signal into the computer and select word boundaries;
- Allocation of parameters characterizing the signal spectrum;
- The use of artificial neural network to evaluate the degree of proximity of acoustic parameters;
- Comparison with standards in the dictionary [8].

Voice signal as an input to a neural network, after processing the audio data received an array of segments of the signal. Each segment corresponds to a set of numbers that characterize the amplitude spectra of a signal, to prepare for the calculation for the signal outputs of the neural network to write all the numbers shows in **Table 2**, where a row which is a set of numbers of each frame.

Where *I* is the number of values of a set of numbers, *N* is the number of sets of numbers (frame signal after slicing). The number of input and output neurons is known, each of the input neurons corresponds to one set of numbers, and the output layer only one neuron, which corresponds to the desired value of the signal recognition. **Table 3** shows the parameter definition uses in this research as shown in **Figure 2**.

4.2. Equations

To calculate the output of the neural network, it's a must complete the following successive steps [9]:

Step 1: Initiate all contexts of all the neurons in the hidden layer;

Step 2: Apply the first set of numbers to the neural network. Calculate the output of the hidden layer.

$$y_j = f\left(\sum_{i=1}^{1} \omega_{ij} X_{1i} + \beta_i + \omega_j X_j \right) \tag{1}$$

F(x)—non-linear activation function

$$y_j = \frac{1}{1 + e^{-\alpha S_j}} \tag{2}$$

for the numbers from 0 to 9.

To recognize the one number you need to build your own neural network it's a must to build 10 of neural networks. Database of over 250 words (numbers from 0 to 9) with different variations of pronunciation, base randomly divided into two equal parts-tutorial and sample tests. When training neural network recognition of one number, for number 5, the desired output of the neural network needs to be unit for the training set with the number 5 and the remainder is zero.
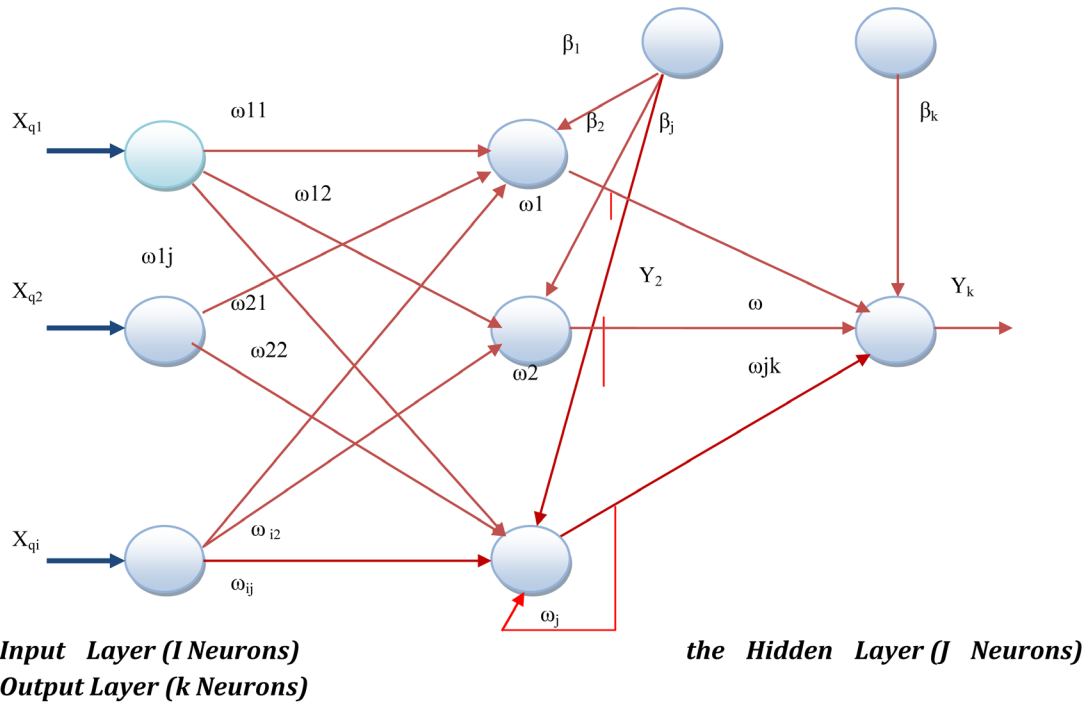
Input Layer (I Neurons)
Output Layer (k Neurons)

the Hidden Layer (J Neurons)

Figure 2. The structure of a neural network with a feedback.

Table 2. Description of a set of speech signal.

Frame	1-value	2-value	…	I-value
1-Frame	x_{11}	x_{12}	…	x_{1I}
2-Frame	x_{21}	x_{22}	…	x_{2I}
…	…	…	…	…
N-Frame	x_{N1}	x_{N2}	…	x_{NI}

Table 3. Parameters definition.

Name	Definition
x_{qi}	*i*-th *q* is the input value to a set of numbers
y_j	Output *j*-neuron layer
ω_{ij}	The weight of the link connecting the *i*-th neuron with the *j*-th neuron
ω_j	weight feedback
β_j	Weight feedback *j*-th neuron; the offset of the *j*-the neuron layer

Neural network training is carried out through the consistent presentation of the training set, with simultaneous tuning scales in accordance with a specific procedure, until around the variety of configuration error reaches an acceptable level. Error in the system function will be calculated by the following formula:

$$E = \frac{1}{2N} \sum_{i=1}^{N} (y_{ki} - d_i)^2 \tag{3}$$

where *N* is the number of training samples processed by neural network examples the real output of the neural network.

A prototype of a neuron is nerve cell biology. A neuron consists of a cell body, or soma, and two types of external wood-like branches: Axon and dendrites. The cell body contains the nucleus, which holds information on hereditary characteristics and plasma with molecular tools for the production and transmission of elements of the neuron of the necessary materials. A neuron receives signals from other neurons through the dendrites and transmits signals generated by the cells of the body, along the axon, which at the end of branches into the fiber, the endings of synapses [1] [3].

Mathematical model of a neuron described democratic ratio:

$$y = f(s), swx_i w_i wb \qquad (4)$$

where w_i is the synapse, the weight (*b*)-offset value, s is the input signal, *y*-signal output neuron, n is the number of inputs to the neuron, *f*-function is activated. Technical model of a neuron is represented in **Figure 3**.

Block diagram of a neuron: $x_1, x_2, \cdots, x_n$-input neuron; $w_1, w_2, \cdots,$ the W_n-a set of weights; $F(S)$ is a function of activation; y-output signal, neuro control performs simple operations like weighted summation, treating the result of nonlinear threshold conversion. Feature of neural network approach is that the structure of the simple homogeneous elements allows you to meet the challenges of the complex relationships between items. The structure of relations defines the functional properties of the network as a whole.

The functional features of neurons and how they combine into a network structure determines the features of neural networks. To meet the challenges of the most adequate identification and management are multilayer neural networks direct action or layered perceptions. When designing neurons together in layers, each of which handles vector signals from the previous layer. Minimum implementation is smiling two-layer neural network, consisting of the input (switch gear), intermediate (hidden), and the output layer [10] (**Figure 4**).
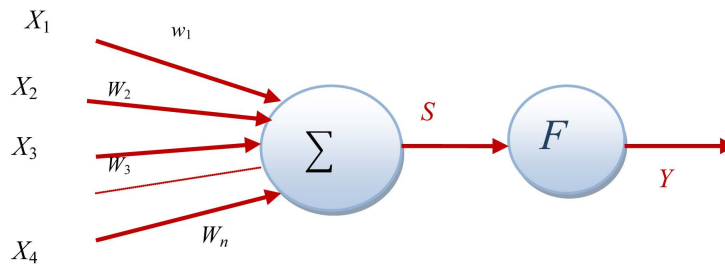

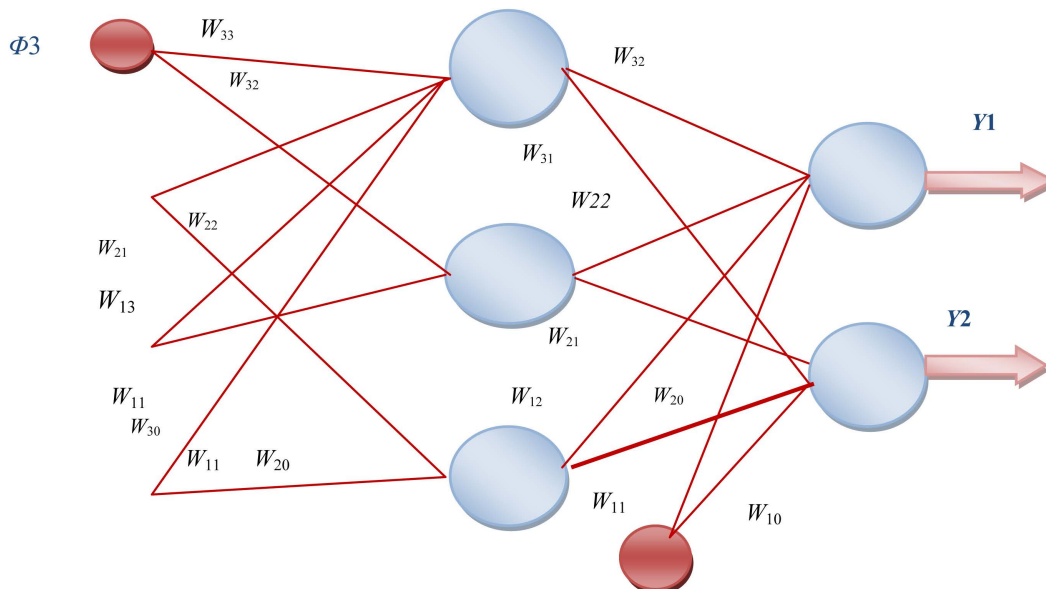**Figure 3.** Technical model of a neuron is represented.

Figure 4. Structural diagram of two-layer neural network.

Implementation of the model of two-layer neural network of direct action has the following mathematical representation:

$$y(\theta) = F_i\left(\sum_{j-1}^{nh} W_{ij} f_i\left(\sum_{j-1}^{nh} w_{ij}\phi_j + w_{jo}\right) + W_{jo}\right) \qquad (5)$$

where the dimension of the vector inputs is: $n\varphi$ φ neural network; nh-the number of neurons in the hidden layer; θ-vector of the configurable parameters of the neural network, which includes weights and neuron-by offset (w_{ji}, W_{ij}); $f_j(x)$-activation function for the hidden layer neurons; $F_i(x)$-activation function neuron in the output layer.

The most important feature of neural network method is the possibility of parallel processing. This feature if there are a large number of international neural connections enables to significantly accelerate the process of signet-data processing [6]. A possibility of processing of speech signals in real time. The neural network has qualities that are inherent in the so-called artificial intelligence [11].

5. Conclusion

Model of speech recognition was based on artificial neural networks. This was investigated to develop a learning neural network using genetic algorithm. This approach was implemented in the system identification numbers, coming to the realization of the system of recognition of voice commands. A system of automatic recognition of speech keywords that were associated with the processing of telephone calls or a sphere of security was developed. The accuracy level of forecasting on the basis of present data set experience was always better.

References

[1] Childer, D.G. (2004) The Matlab Speech Processing and Synthesis Toolbox. Photocopy Edition, Tsinghua University Press, Beijing, 45-51.

[2] Chien, J.T. (2005) Predictive Hidden Markov Model Selection for Speech Recognition. *IEEE Transaction on Speech and Audio Processing*, **13**.

[3] Luger, G. and Stubblefield, W. (2004) Artificial Intelligence: Structures and Strategies for Complex Problem Solving. 5th Edition, The Benjamin/Cummings Publishing Company, Inc.
http://www.cs.unm.edu/~luger/ai-final/tocfull.htm

[4] Choudhary, A. and Kshirsagar, R. (2012) Process Speech Recognition System Using Artificial Intelligence Technique. *International Journal of Soft Computing and Engineering* (*IJSCE*), **2**.

[5] Ovchinnikov, P.E. (2005) Multilayer Perceptron Training without Word Segmentation for Phoneme Recognition. *Optical Memory & Neural Networks* (*Information Optics*), **14**, 245-248.

[6] Guo, X.Y., Liang, X. and Li, X. (2007) A Stock Pattern Recognition Algorithm Based on Neural Networks. *Third International Conference on Natural Computation*, **2**.

[7] Dai, W.J. and Wang, P. (2007) Application of Pattern Recognition and Artificial Neural Network to Load Forecasting in Electric Power System. *Third International Conference on Natural Computation*, **1**.

[8] Shahrin, A.N., Omar, N., Jumari, K.F. and Khalid, M. (2007) Face Detecting Using Artificial Neural Networks Approach. *First Asia International Conference on Modelling & Simulation*.

[9] Lin, H., Hou, W.S., Zhen, X.L. and Peng, C.L. (2006) Recognition of ECG Patterns Using Artificial Neural Network. *Sixth International Conference on Intelligent Systems Design and Applications*, **2**.

[10] Al Smadi, T.A. (2013) Design and Implementation of Double Base Integer Encoder of Term Metrical to Direct Binary. *Journal of Signal and Information Processing*, **4**, 370.

[11] Takialddin Al Smadi Int. An Improved Real-Time Speech Signal in Case of Isolated Word Recognition. *Journal of Engineering Research and Applications*, **3**, 1748-1754.